DATA607 - Homework 1

Mike Dehn

9/4/2021

1 Effective and Ineffective Visualizations

1.1 Effective Visualizations

The following subsections provide examples of effective graphical displays of data.

1.1.1 Example 1

1.1.2 Example 2

1.2 Ineffective Visualizations

The following subsections provide examples of ineffective graphical displays of data.

1.2.1 Example 1

This example is taken from the FleetLogging, Inc. study of which subway stations are the most stressful for travelers. The study used TensiStrength to analyze stress levels in the text of Google reviews for subway stations across the globe. Researchers built reports on the relative stress levels experienced by travelers in each subway system and provided a ranking for each station. Figure 1 represents the study's findings in New York City.

The FleetLogging figure is ineffective from a graphical design perspective. The data presented are the stress percentage levels for each subway station in the NYC subway system and could be much better represented using a simple bar chart. This figure is technically a bar chart that wraps around in a circular shape. It is difficult to tell which percentage values are associated with each station since the eye must follow a circular path from the name to the percentage. A simple horizontal or vertical bar chart is much easier to understand.

The format of the figure is also misleading and might cause confusion about the study's findings. Since each bar in the chart is set at a different radius from a concentric circle, the overall area of each bar changes dramatically based on its position in the chart. The minimum stress value is 53.7 percent (Sheepshead Bay) and the maximum stress value is 66.7 percent (Jamaica Center-Parsons/Archer). Although the maximum value is approximately 24 percent larger than the minimum, its bar area is many times larger than that of the minimum. This may cause readers to misinterpret the difference in stress levels as more dramatic than they actually are.

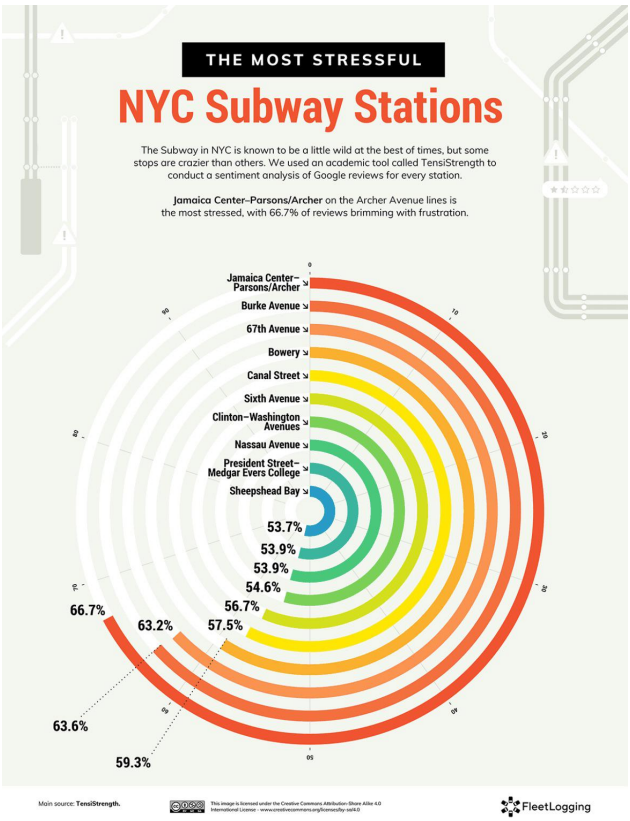

Figure 1: New York City subway station stress rankings based on TensiStrength analysis of Google reviews

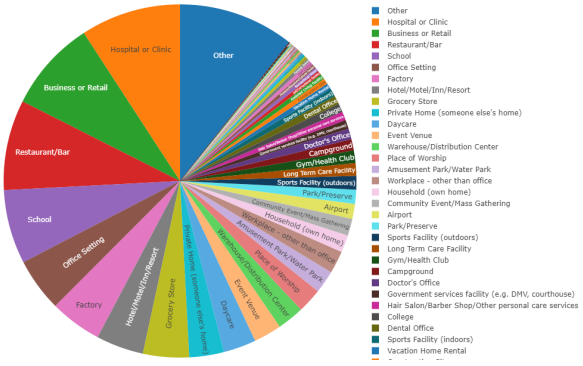

Figure 2: Prevalence of COVID case exposure sources

1.2.2 Example 2

This example is taken from the Illinois Department of Public Health study on COVID-19 case exposure locations. The data in this study were gathered from contact tracing of COVID-19 patients. Figure 2 represents the relative proportion of case exposure locations.

The COVID-19 figure is ineffective from a graphical design and informational perspective. Although a pie chart is a good choice for showing relative proportions of the values for a one-dimensional data set, the chart is unreadable because the chart is divided into over 50 sections. The percentage values are unreadable for most of the categories in data and would benefit from a less granular summary-level view of location categories.The problem is exacerbated by the original source not providing another format (e.g., a table) that contains these values. In essence, there is information reported, but lost, due to this ineffective graphic.

The chart may also be misleading since there is no "Unknown" category in the data. Some patients likely responded to the contact tracing questionnaire with an unknown exposure location, but this is not reported as a distinct category from "Other". As a result, the audience cannot interpret the difference between unexplained exposures and those that come from a known but non-specific category. # Part 2